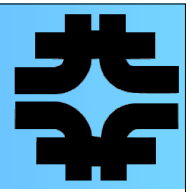


Storage Services and UAF Status/Plans

June 14, 2004

Jon Bakken

Storage Services

CMS Facilities major collaboration with CCF and CSS departments

- CCF - storage systems, networking, file transfers
- CSS - system administration, core services, cluster adminstration
-
-

CMS has a 4 Tier Grid Architecture

- CERN is Tier 0, place where data is created
- FNAL is Tier 1, North American Regional Center
- Project supported at 5 Big Universities are Tier 2
- Local analysis clusters elsewhere are Tier 3

Storage Services

Data Transfer Protocol will be GridFTP

- Need many extensions described in extensions of standard protocol, described in detail in http://www.gridforum.org/Meetings/GGF11/Documents/draft-ggf-gridftp-v2protocol-1.pdf
  - IVM leading this effort, just approved at GGF
  - Contained in new mode "X-mode",
  - Protocol remains compatible with previous versions
  -
- CRC of data blocks or files for reliable transfers
-
- Eliminate PASV defect - need to know address before filename
  - Finally will provide scalable distributed file transfers
-
- Eliminate firewall issues - clients can be active for all transfers

Storage Services

SRM will be management protocol and provides uniform access to data

TP leading SRM development

- Provides for grid reservation and scheduling of storage resources
  - Need database improvements for reliable transfers (transient failures)
  - Need new scheduler algorithm - new based on experience (retries)
  - Need space reservation
  -

SRM should work with any storage element.

- Need SRM to be separate product capable of being used with a wide variety of storage elements (dCache, NeST, sam-cache)

Storage Space

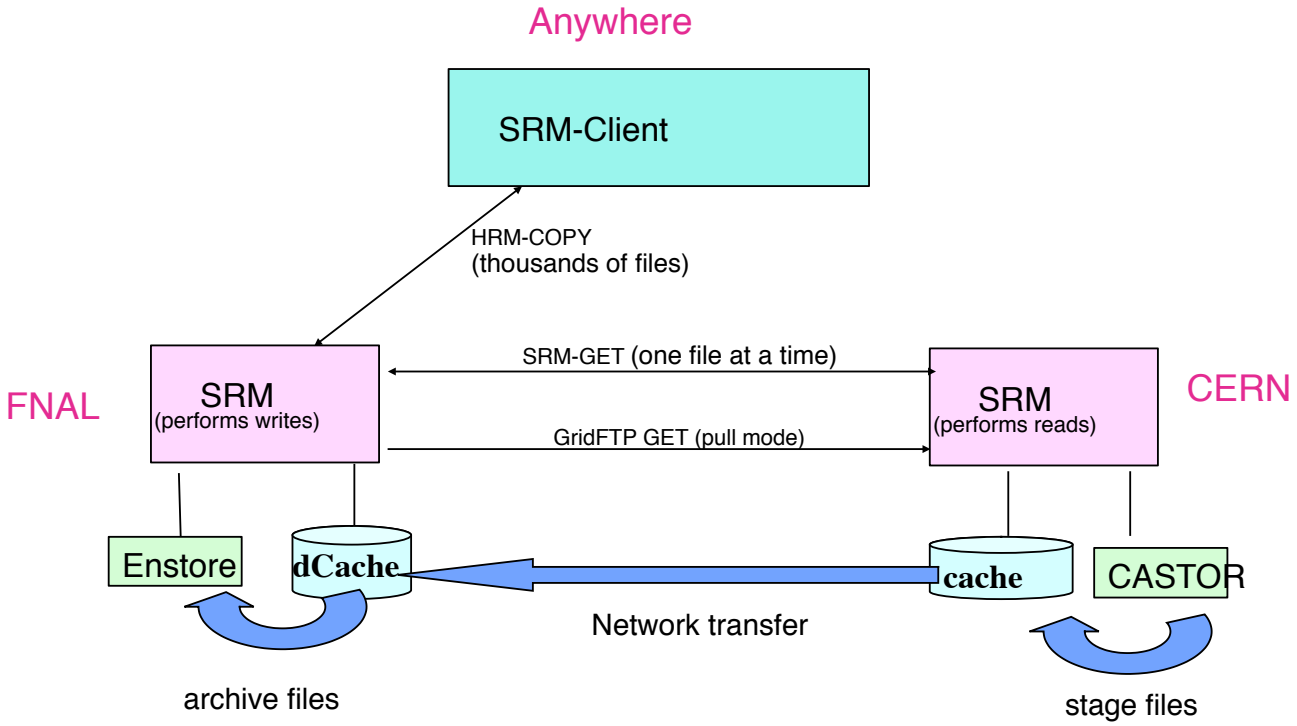
- Expect to need Permanent, durable (limited lifetime, not archived) and Volatile (durable + can be auto-deleted after lifetime) storage spaces

Storage Services

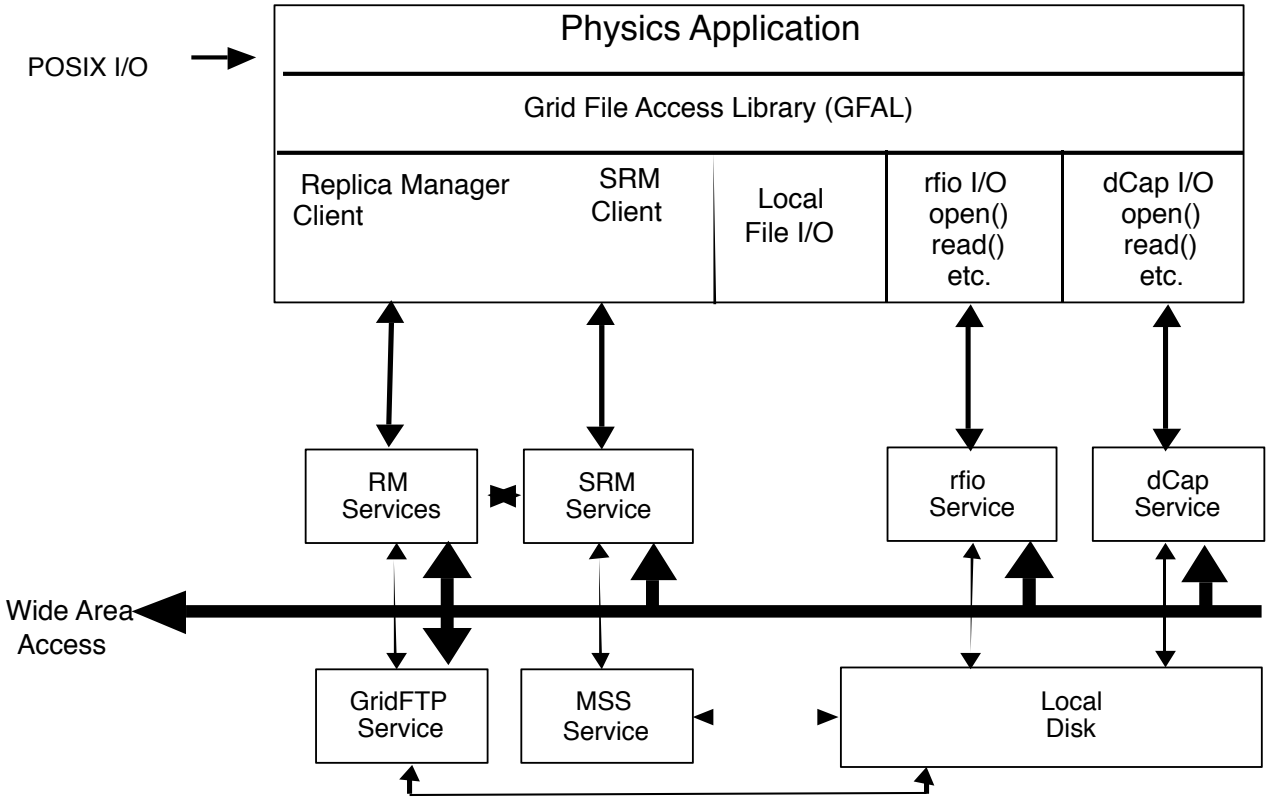
Currently have about 20 TB of dCache disk pools at FNAL

- Roughly 10 TB read pools, 10 TB write pools

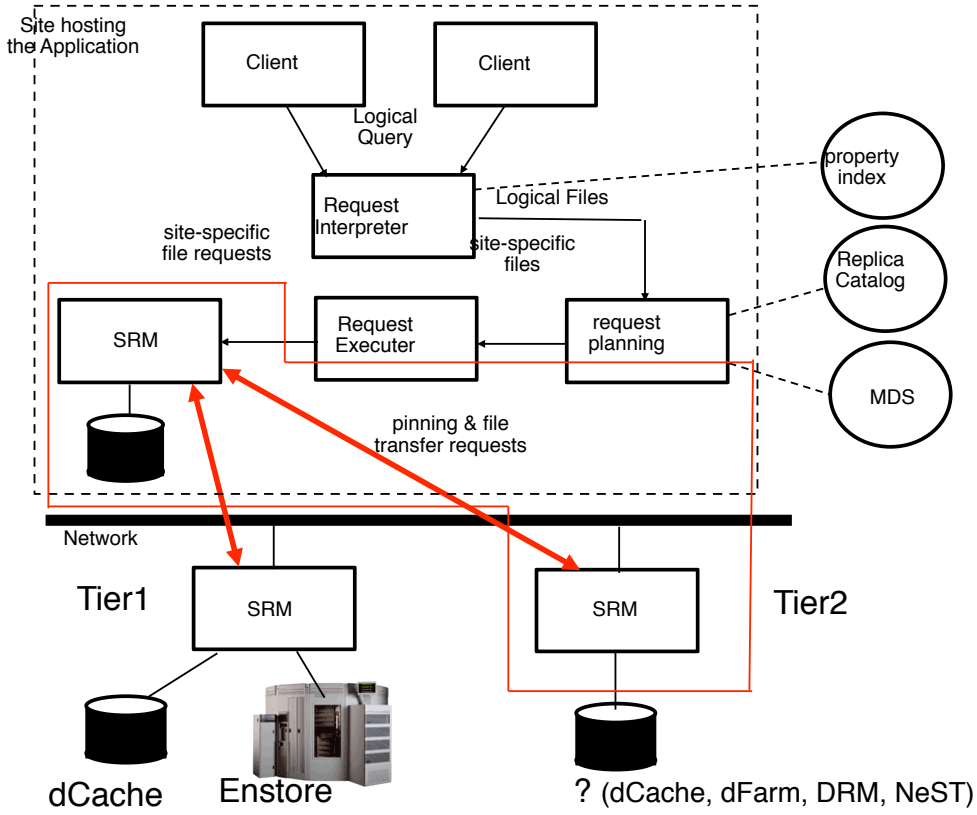
100 TB permanent tape archive



Storage Services



Storage Services



Storage Services

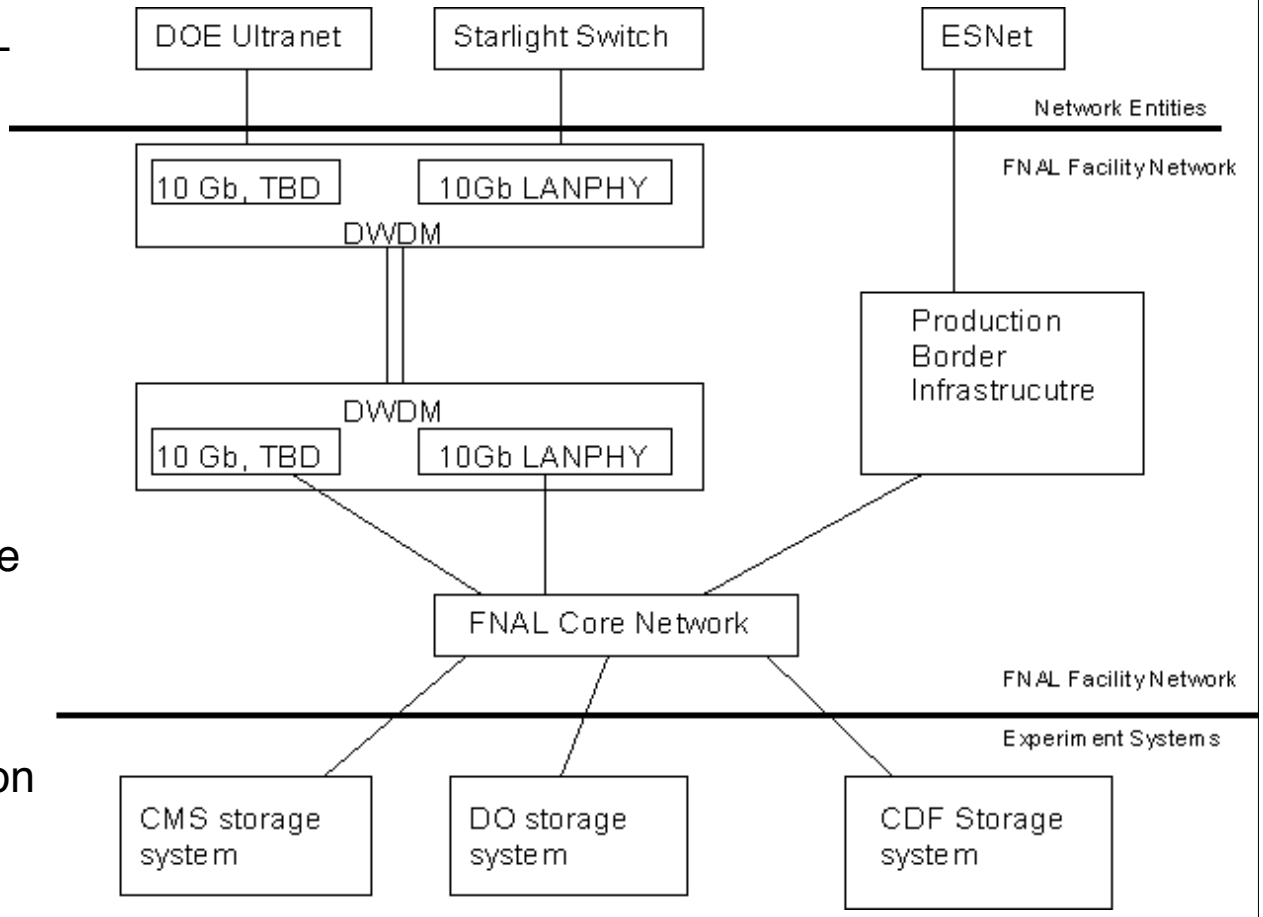
Optimizing WAN Transfers - WAWG effort

Starlight path bypasses FNAL border router

Initially, static routes between src/dst pairs, but goal is aggregation of many flows to fill a (dynamic) pipe, “owned” by CMS VO.

Forwarding to the pipe is done on a per flow basis

Starlight path ties directly to production LAN and production Storage Element (no dual NICs).



Storage Services

Robust Transfer Challenge, CERN to Tier 1 sites

- Goal is not amount of data, but reliability of service
- Problems meant to be fixed, not just restarts/reboots

- Goal is reliable service at 100 MB/s for 1 week for each site
  - Need to use Starlight to achieve this (can not use all of it)

- After individual sites robust, 500 MB/s for CERN to 5 sites at once

- Coordination started at HEPIX, just starting

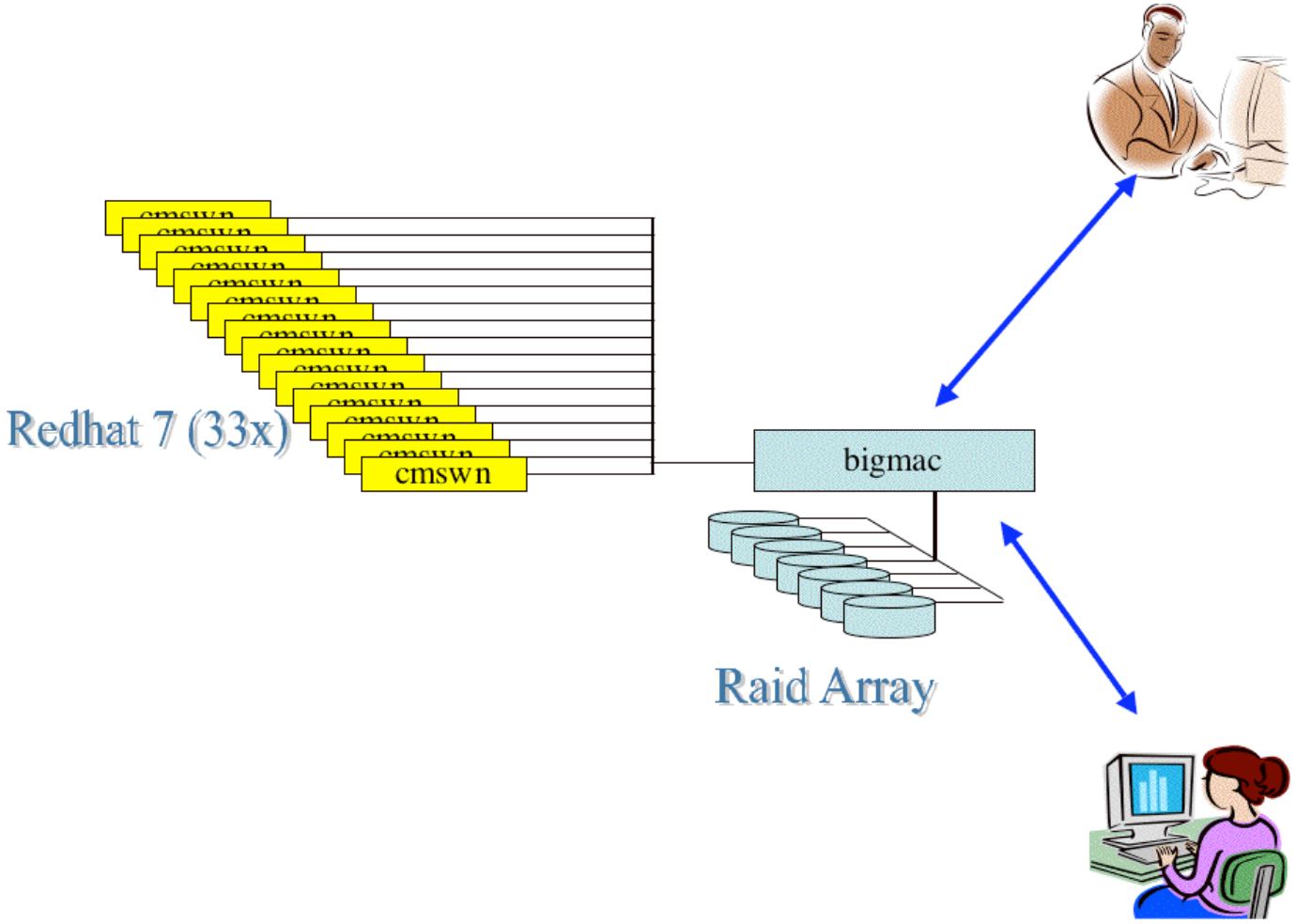
User Analysis Facilities

The FNAL User Analysis Facilities are part of Tier 1 Facilities that provide local interactive and batch computing for users

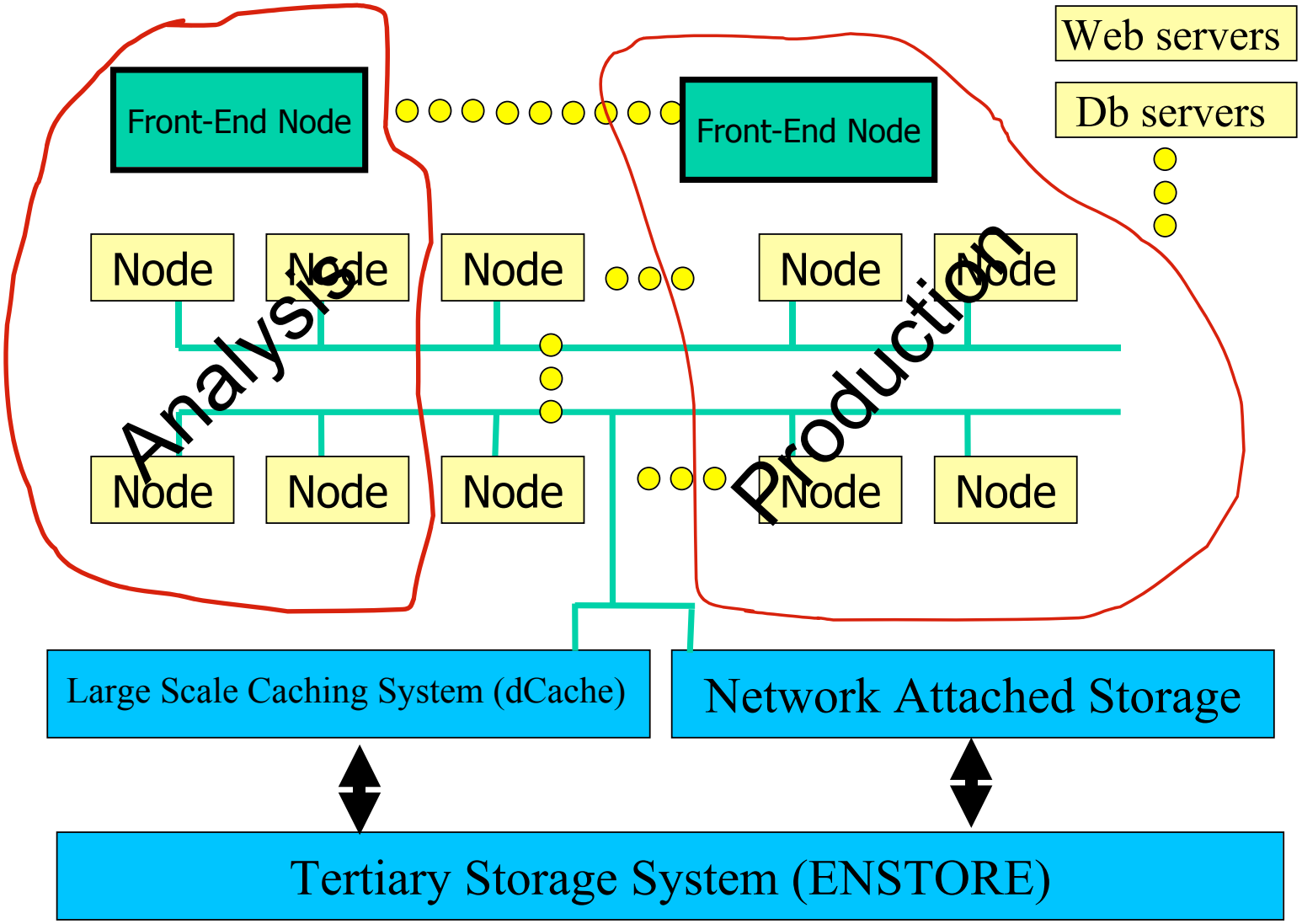
- Uniform software configuration management of CMS software
  - CMS simulation, including pileup
- 160 users - used for Monte Carlo production by FNAL physics groups, including the LPC tracking group and jetmet physics group….
  - Several tutorials were well received (more than 30 people at each)
- Load balancing via FBSNG, investigating more standard solutions
- High performance data service
  - dCache and Enstore access to data
  - IBRIX distributed file system - connected over a fiber channel switch
    - All worker nodes connected over gigabit using interconnected small switches
  - External data exchange service (CMSXFER)
    - Currently only scp, extending to gridftp soon

User Analysis Facilities

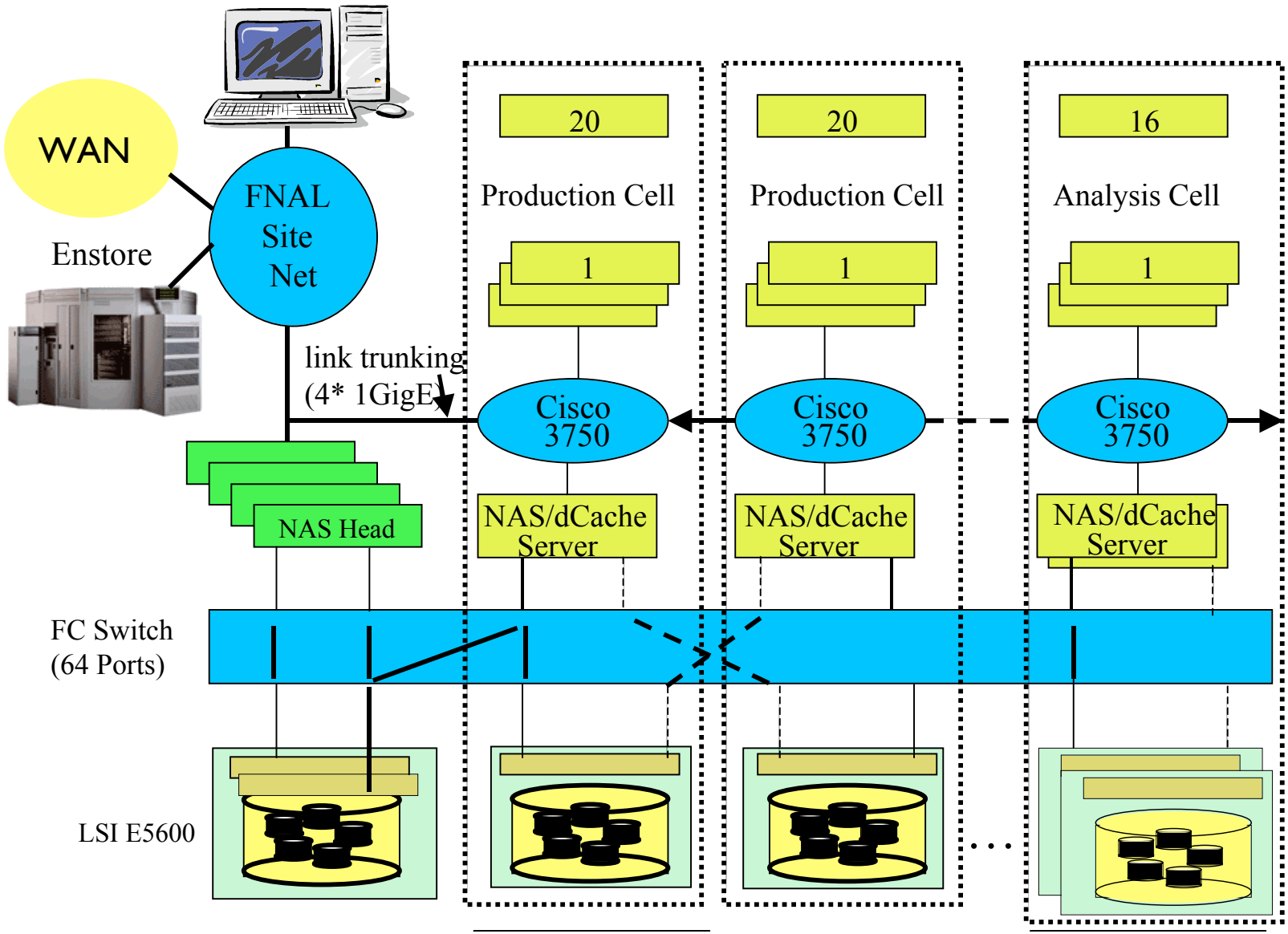
http://www.uscms.org/scpages/general/users/farm/CMS_UAF.html



ssh -t bigmac.fnal.gov *(+r RH7)*


Web servers
Db servers
Front-End Node
Front-End Node
Node
Node
Node
Node
Node
Node
Node
Node
Node
Analysis
Production
Large Scale Caching System (dCache)
Network Attached Storage
Tertiary Storage System (ENSTORE)

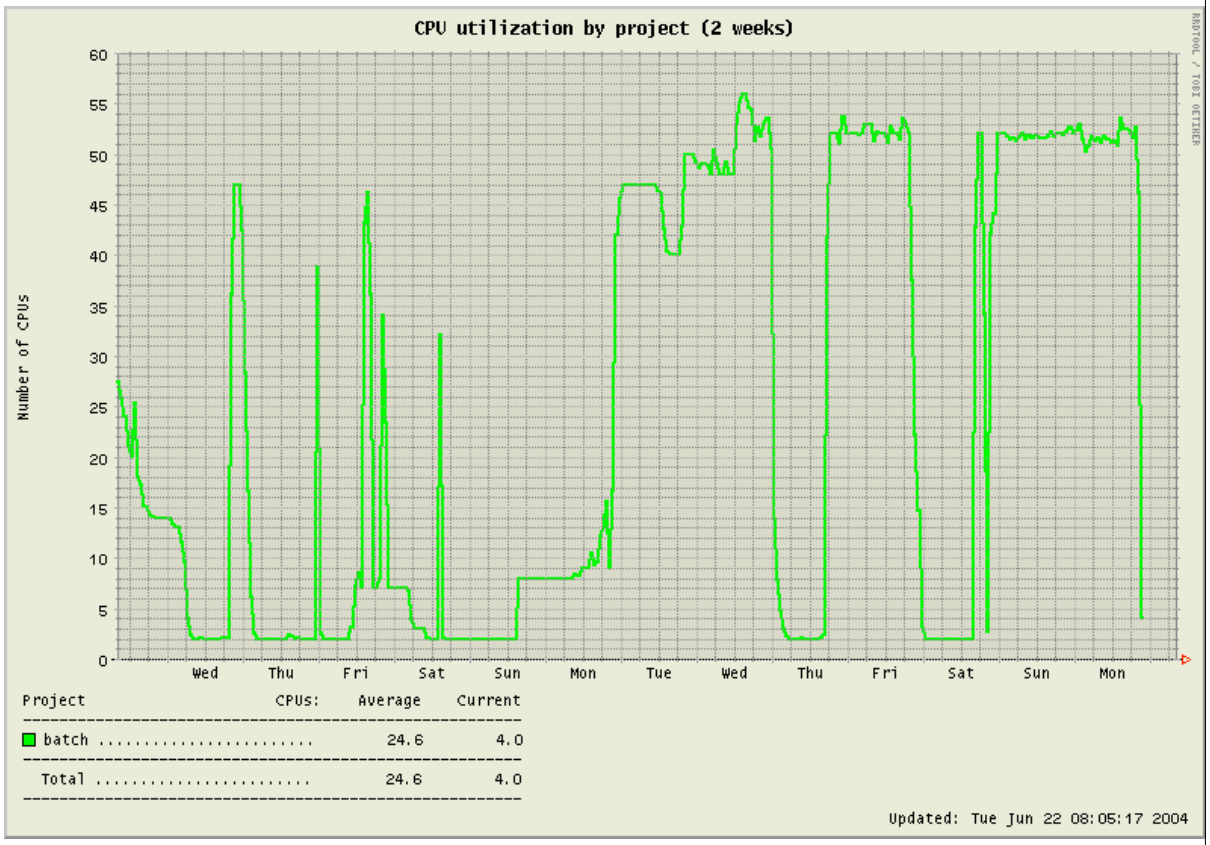
User Analysis Facilities



User Analysis Facilities

FBSNG queues

- Interactive - 24 hr limit, 4 per node
- Long - no time limit
- Medium - 8 hour max
- Short - 1 hour max



Non-interactive usage, 55 slots available

UAF Improvements

- Adding more IBRIX disk space
- Home areas - investigating using IBRIX instead of AFS
- Waiting for General CD backup solution
- Investigating load balancing (bigmac is single point of failure)